MATH 365 Stochastic Processes

Goals: (1) gain some intuition about stochasticity & randomness
(2) learn to simulate & analyze random processes

We'll begin with finite-state, discrete time processes, modeled as a sequence of random variables $X_0, X_1, X_2, \cdots$

Example: Bernoulli process X_n = result of n^{th} coin flip
(1 for heads, 0 for tails)
(R demo)
State space Ω = all binary sequences

Example: Simple random walk $S_n = X_1 + X_2 + \cdots + X_n$, $X_k = \begin{cases} 1 & \text{heads} \\ -1 & \text{tails} \end{cases}$
(R demo)
Heads $\to$ move up
Tails $\to$ move down

Markov chain: X_{n+1} depends on X_n but no other previous X_k

Knowing $\mathbb{P}\{X_{n+1} = j \mid X_n = i\}$ for all states i & j, $n \geq 0$, characterizes the MC

Brief probability review

- State space Ω = set of all outcomes of a random experiment (possible states the stochastic process can take)

Event A = subset of Ω

Conditional probability of event A occurring given that event B occurred

$$\mathbb{P}(A \mid B) = P(A \cap B) / P(B)$$

A & B are <u>independent</u> if $P(A \mid B) = P(A)$

Law of Total Probability: If $B_1, \cdots, B_k$ are disjoint events and their union is all of Ω, then for any event A

$$\mathbb{P}(A) = \sum_{i=1}^{k} \mathbb{P}(A \cap B_i) = \sum_{i=1}^{k} P(A \mid B_i) P(B_i)$$

→ Example: 7% of men are colorblind and 0.4% of women are colorblind. 49% of population is male, 51% female.
What is prob a randomly selected person is colorblind?

$$\mathbb{P}(\text{colorblind}) = P(C \mid M)P(M) + P(C \mid F)P(F) = (.07)(.49) + (.004)(.51)$$
$$= .03634$$

Bayes Rule $\mathbb{P}(B|A) = \dfrac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)}$ ← $\mathbb{P}(A)$

complement
(B didn't occur)

→ Example: Suppose one person in a 100-person company is a thief.
Company administers a polygraph test to its employees.
If subject is a liar, there is 95% chance polygraph will detect this.
If subject is truthful, there is 10% chance of false positive (liar).
An employee is chosen at random and given polygraph test,
says they are not a thief, but polygraph says liar.
What is prob that employee is lying?

$$\mathbb{P}(\text{Liar} \mid \text{Detected lie}) = \frac{\mathbb{P}(D|L)\,\mathbb{P}(L)}{\mathbb{P}(D|L)\mathbb{P}(L) + \mathbb{P}(D|L^c)\,\mathbb{P}(L^c)}$$

$$= \frac{(.95)(.01)}{(.95)(.01) + (.10)(.99)}$$

$$= .088$$ very low!

Expected value $\mathbb{E}[X] = \sum_{k \in \Omega} k \, \mathbb{P}(X=k)$

Conditional expectation $\mathbb{E}[Y \mid X=x] = \sum_{y} y \, \mathbb{P}(Y=y \mid X=x)$

Law of Total Expectation: Let Y be a random variable and $A_1, \cdots, A_k$ a partition of the state space Ω.

Then $\mathbb{E}[Y] = \sum_{i=1}^{k} \mathbb{E}[Y \mid A_i] \mathbb{P}(A_i)$

Example: Flip fair coin repeatedly. What is the expected # of flips to get 2 heads in a row?

Let Y = # flips

$\mathbb{E}[Y \mid 1^{st} \text{ flip is } T] = 1 + \mathbb{E}[Y]$ (1st flip no good, like starting over on 2nd flip)

$\mathbb{E}[Y \mid 1^{st} \text{ 2 flips } HT] = 2 + \mathbb{E}[Y]$ (1st 2 flips no good)

$\mathbb{E}[Y \mid 1^{st} \text{ 2 flips } HH] = 2$ (done!)

$\Rightarrow \mathbb{E}[Y] = (1 + \mathbb{E}[Y])(.5) + (2 + \mathbb{E}[Y])(.25) + 2(.25)$

Solve to get $\mathbb{E}[Y] = 6$

Law of Total Expectation $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid X]]$

Proof: $\mathbb{E}[\underbrace{\mathbb{E}[Y \mid X]}] = \sum_x \mathbb{E}[Y \mid X = x]\, \mathbb{P}(X = x)$

treat as RV, depends on value X takes

$$= \sum_x \left(\sum_y y\, \mathbb{P}(Y = y \mid X = x) \right) \mathbb{P}(X = x)$$

$$= \sum_y y \left(\sum_x \mathbb{P}(Y = y \mid X = x)\, \mathbb{P}(X = x) \right)$$

(assuming interchange of sums is valid)

$$= \sum_y y\, \mathbb{P}(Y = y)$$

$$= \mathbb{E}[Y]$$

→ Example: Random sum of random variables

Let $T = X_1 + \cdots + X_N$, where N is also a RV & X_k iid RVs

iid = independent identically distributed

$$\mathbb{E}[T] = \mathbb{E}[\mathbb{E}[T \mid N]] = \mathbb{E}[N\, \mathbb{E}[X]]$$

$$= \mathbb{E}[N]\, \mathbb{E}[X]$$ Wald's Eq

Sections 2.1–2.2 Markov chains

Let Ω be a discrete set of states (can be finite or infinite)

A <u>Markov chain</u> is a sequence of RVs $X_0, X_1, X_2, \cdots$ such that

$$\mathbb{P}(X_{n+1}=j \mid X_0=x_0, X_1=x_1, \cdots, X_{n-1}=x_{n-1}, X_n=i)$$

$$=\mathbb{P}(X_{n+1}=j \mid X_n=i) \quad \text{for all states and } n\geq 0$$

That is, only X_n affects X_{n+1}, no other past values matter.

We can organize these probabilities into a <u>transition matrix</u> P

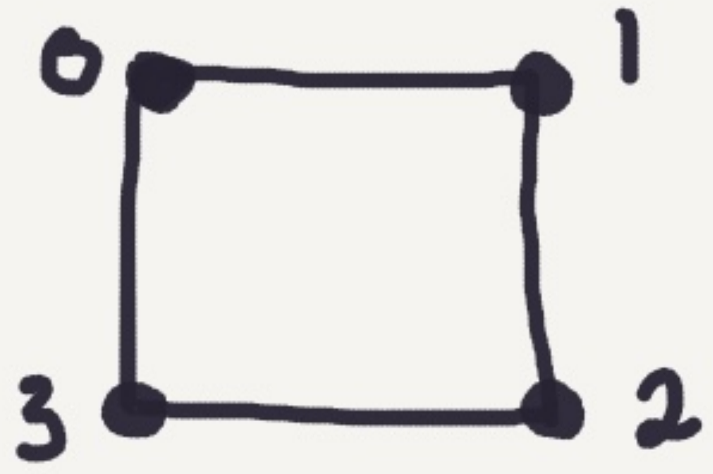
with entries $p_{ij}=\mathbb{P}(X_{n+1}=j \mid X_n=i)$

"one-step transition probabilities"

(probability of going from state i to j in one step)

Each row of P will sum to one, and all entries are nonnegative.
Such a matrix is called a "stochastic matrix."

Example: simple random walk on a graph



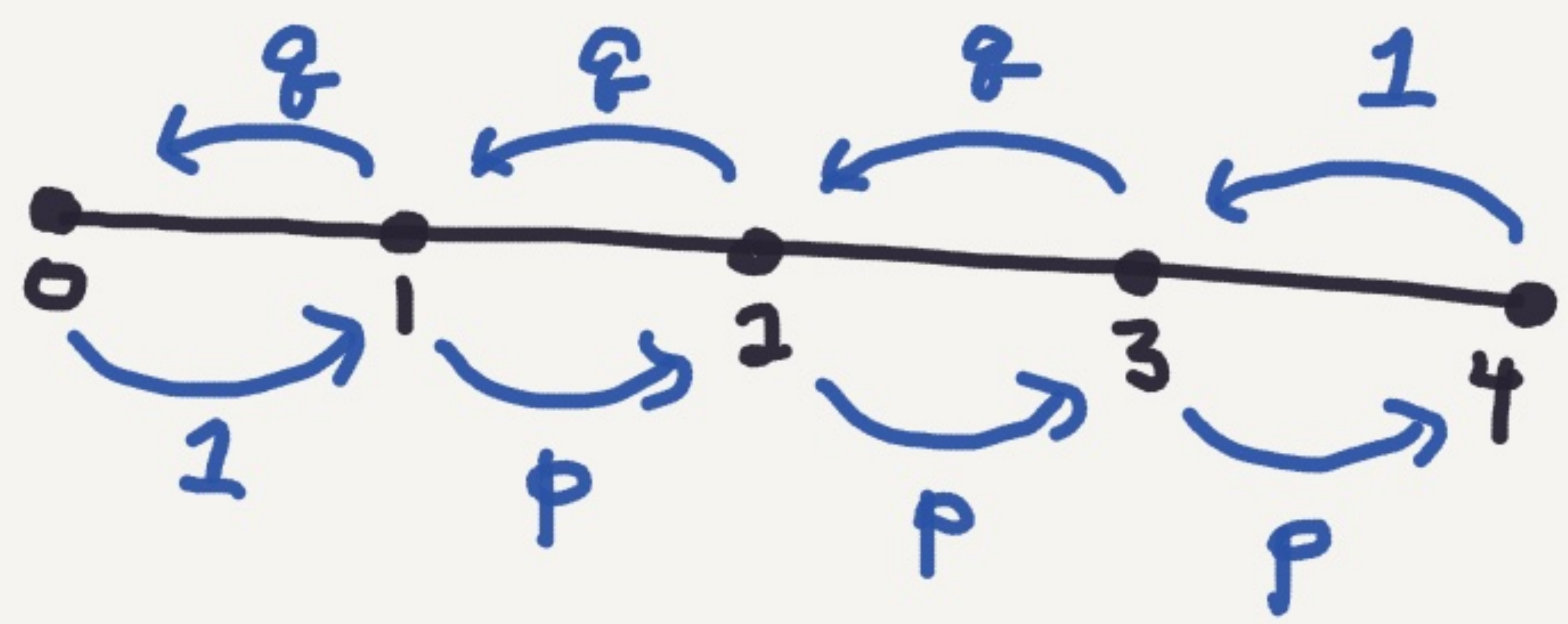
equally likely to follow any edge connected to current vertex

(for example, if currently at node 2, 50-50 chance of going to node 1 or 3)

Transition matrix

$$P = \begin{array}{cc} & \text{to} \\ & \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \text{from} \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{bmatrix} \end{array}$$

Example: Random walk with reflecting boundaries

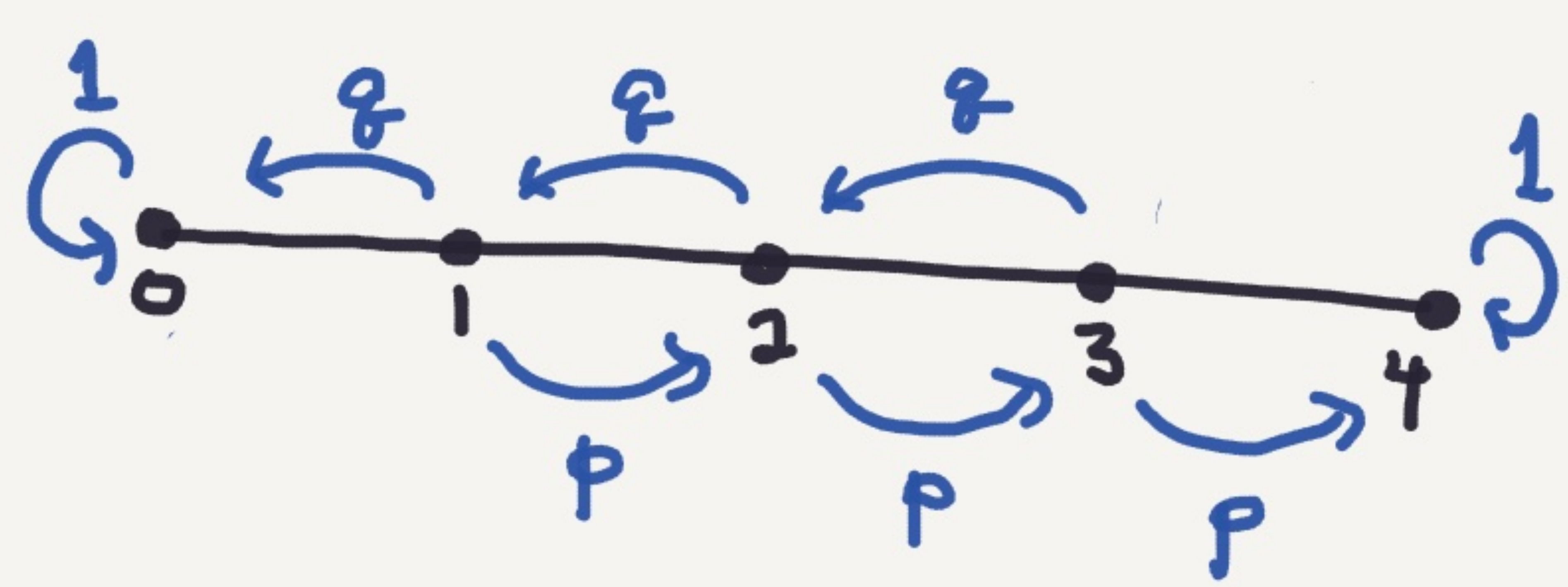


If at state 1, 2, or 3, flip a coin to determine whether to move right with probability p or left with probability $q = 1 - p$

If at state 0, go next to 1
If at state 4, go next to 3

$$\text{Transition matrix } P = \begin{array}{c c} & \text{to} \\ & \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \text{from} \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ q & 0 & p & 0 & 0 \\ 0 & q & 0 & p & 0 \\ 0 & 0 & q & 0 & p \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \end{array}$$

R demo: Gambler's ruin
absorbing BCs



Probability vector: components nonnegative & sum to one
(gives probability of being in each state)

Initial prob vector: indicates how to start Markov chain
(prob distribution for X_0)

- can specify a particular state

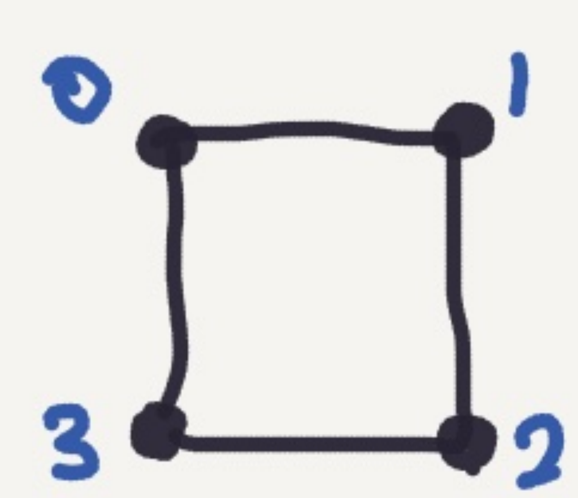
  $\alpha = [\,0 \;\; 0 \;\; 1 \;\; 0 \;\; 0\,]$

- can be a starting distribution with prob of beginning chain in each state

  $\alpha = [\,0 \;\; \tfrac{1}{2} \;\; 0 \;\; \tfrac{1}{2} \;\; 0\,]$

Multiplying α by P gives probability distribution for X_1

$$\overset{\text{start at}}{\begin{array}{cccc} 0 & 1 & 2 & 3 \\ \hline \end{array}}\;[\,0 \;\; 0 \;\; 1 \;\; 0\,]\;\; \begin{array}{c|cccc} \text{from} \backslash \text{to} & 0 & 1 & 2 & 3 \\ \hline 0 & 0 & \tfrac{1}{2} & 0 & \tfrac{1}{2} \\ 1 & \tfrac{1}{2} & 0 & \tfrac{1}{2} & 0 \\ 2 & 0 & \tfrac{1}{2} & 0 & \tfrac{1}{2} \\ 3 & \tfrac{1}{2} & 0 & \tfrac{1}{2} & 0 \end{array} \;=\; \begin{array}{cccc} 0 & 1 & 2 & 3 \\ \hline 0 & \tfrac{1}{2} & 0 & \tfrac{1}{2} \end{array}$$



Section 2.3

n-step transition probabilities

$$p_n(i,j) = \mathbb{P}(X_n = j \mid X_0 = i) = \mathbb{P}(X_{n+k} = j \mid X_k = i) \quad \text{for all } n, k \geq 0$$

(probability of going from state i to state j in exactly n steps)

$$\mathbb{P}(X_{n+1} = j \mid X_0 = i) = \sum_{k \in \Omega} \mathbb{P}(X_n = k \mid X_0 = i)\,\mathbb{P}(X_{n+1} = j \mid X_n = k)$$

all ways of getting to state j at time $n+1$

$$= \sum_k p_n(i,k)\,p(k,j)$$

matrix multiplication formula!

Using induction, we see $P_{n+1} = P_n P = P^n P = P^{n+1}$

where P_n has entries $p_n(i,j)$ (that is, $P_n = P^n$ for all $n \geq 0$)

If the initial prob vector is α, then the probability distribution for X_n is αP^n.

Chapman-Kolmogorov relationship

$$P_{m+n} = P_m P_n = P^m P^n \quad \text{(generalizes previous equation)}$$

Written out in terms of probabilities,

$$\mathbb{P}(X_{m+n}=j \mid X_0=i) = \sum_k \mathbb{P}(X_n=k \mid X_0=i)\,\mathbb{P}(X_{m+n}=j \mid X_n=k)$$

$$= \sum_k \mathbb{P}(X_n=k \mid X_0=i)\,\mathbb{P}(X_m=j \mid X_0=k)$$

by Markov property (transition probs are same at all steps)

$$= \sum_k p_n(i,k)\,p_m(k,j)$$

R demo: matrix powers

Example: Markov chain with state space $\{0, 1\}$

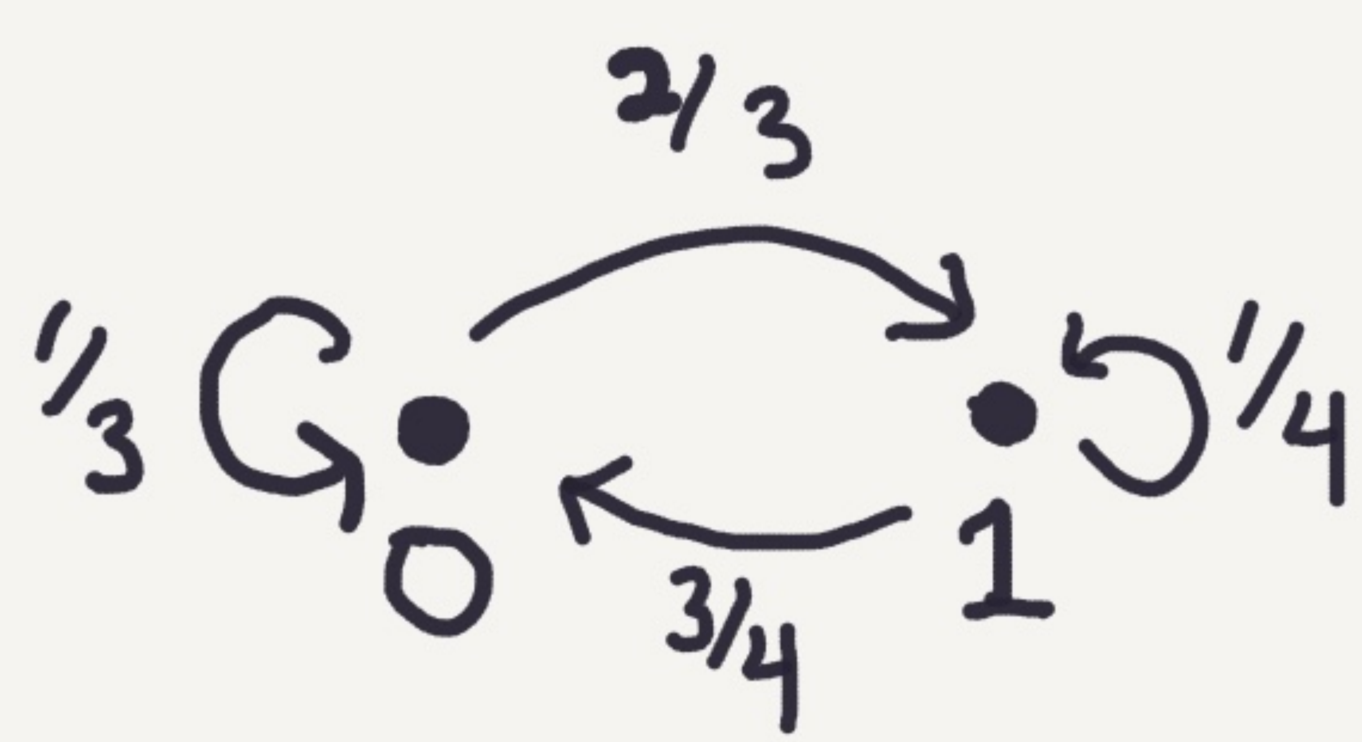
and transition matrix $P = \begin{array}{c|cc} & 0 & 1 \\ \hline 0 & & \\ 1 & & \end{array}$ $\begin{bmatrix} 1/3 & 2/3 \\ 3/4 & 1/4 \end{bmatrix}$

Assume initial prob vector $\alpha = \begin{array}{cc} 0 & 1 \\ \end{array}$ $[\,1 \quad 0\,]$ (always start in state 0)

What is prob of being in state 1 at $n = 3$?

$$\alpha P^3 = [\,1 \quad 0\,] \begin{bmatrix} .495 & .505 \\ .568 & 432 \end{bmatrix} \quad \text{so} \quad \mathbb{P}(X_3 = 1 \mid X_0 = 0) = .505$$

$$= \begin{array}{cc} 0 & 1 \\ \end{array}$$
$$[\,.495 \quad \underline{.505}\,]$$



Section 2.4 Long-term behavior and invariant prob.

What happens over the long run to $\mathbb{P}(X_n = j \mid X_0 = i)$ as $n \to \infty$?

We need to see what happens with P^n as $n \to \infty$.

If $\lim_{n\to\infty} \alpha P^n$ exists (converges to some $\bar{\pi}$ vector),

then we'll call $\bar{\pi}$ the "invariant prob vector"

(steady state, stationary, equilibrium vector)

In this case, does $\bar{\pi}$ depend on α?

R demo: 5 state reflecting boundaries - P^n does not converge

R demo: $P = \begin{bmatrix} 3/4 & 1/4 \\ 1/6 & 5/6 \end{bmatrix} \Rightarrow P^n$ converges to $\begin{bmatrix} .4 & .6 \\ .4 & .6 \end{bmatrix}$

$$\alpha P^n = [\alpha_0 \ \alpha_1] \begin{bmatrix} .4 & .6 \\ .4 & .6 \end{bmatrix} = [.4 \ \ .6]$$

Note $\alpha_0 + \alpha_1 = 1$

Same for all α!

<

Questions:

1) Do all stochastic matrices have at least one invariant prob vector (or at most one)?

2) What characterizes matrices with a <u>unique</u> invariant prob vector $\bar{\pi}$?

3) When is it true that $\lim_{n\to\infty} P^n = \begin{bmatrix} \bar{\pi} \\ \vdots \\ \bar{\pi} \end{bmatrix}$, which implies

$$\lim_{n\to\infty} \alpha P^n = \bar{\pi} \text{ for all initial vectors } \alpha ?$$

Some linear algebra will help us answer these questions.

Question 1: Transition matrix P has rows that sum to one,

so $$P\begin{bmatrix}1\\1\\ \vdots\\1\end{bmatrix} = \begin{bmatrix}p_{11} & p_{12} & \cdots & p_{1n}\\ p_{21} & p_{22} & \cdots & p_{2n}\\ \vdots & & & \\ p_{m1} & p_{m2} & \cdots & p_{mn}\end{bmatrix}\begin{bmatrix}1\\1\\ \vdots\\1\end{bmatrix} = \begin{bmatrix}\text{sum of row 1}\\ \text{sum of row 2}\\ \vdots\\ \text{sum of last row}\end{bmatrix} = \begin{bmatrix}1\\1\\ \vdots\\1\end{bmatrix}$$

(Eigenvalue λ with associated eigenvector $\vec{v}$ satisfies $P\vec{v} = \lambda\vec{v}$. A left eigenvector satisfies $\vec{v}P = \lambda\vec{v}$. Eigenvalues will be same, but left & right eigenvectors differ.)

Hence $\lambda = 1$ is an eigenvalue of P with right eigenvector $\begin{bmatrix}1\\ \vdots\\1\end{bmatrix}$

P and P^t have the same eigenvalues, which implies

P has eigenvalue $\lambda = 1$ for some <u>left</u> eigenvector v : $vP = v$

which is an invariant prob vector.

Answer: every transition matrix P will have at least one invariant prob vector.

Question 2: When is there exactly one invariant prob vector?

R demo: absorbing boundaries example has multiple long-run behaviors, depending on initial state

note $\lambda = 1$ has multiplicity 2, and any linear combination of the associated left eigenvectors is a possible long-run distribution

Thm: If P is a stochastic matrix such that for some $n \geq 1$, P^n has all positive entries (that is, can reach all possible states from any initial vector), then P has eigenvalue $\lambda = 1$ with multiplicity 1, with unique associated left eigenvector π having all positive entries summing to one.

In addition, all other eigenvalues of P have abs value < 1 and $\lim_{n \to \infty} P^n = \begin{bmatrix} \pi \\ \vdots \\ \pi \end{bmatrix}$

The theorem also answered Question 3.

We now have two ways to determine $\bar{\pi}$ when P has $\lambda = 1$ with multiplicity 1

1. Raise matrix to high power to estimate $\lim_{n\to\infty} P^n$ and take any row as $\bar{\pi}$

2. Find the left eigenvector of P for $\lambda = 1$ (normalize so sums to one)

In R:

```
r <- eigen(t(P))
V <- r$vectors
lambda <- r$values
pibar <- V[,1]/sum(V[,1])
```

check which eigenvalue is 1, not always first one

Recap:

* If a stochastic process satisfies the property that the transition probabilities only depend on the current state and no past states, it is a Markov chain (assuming same transition matrix for all time steps).
* The one-step transition matrix P stores the transition prob of going from state i to state j in entry w/row i, column j.
* The n-step transition matrix is P^n, giving the prob of going from state i to state j in n steps.
* Given an initi'al prob vector α, the probability of ending up in state j after n steps is the j^{th} component of αP^n.
* If P has exactly one eigenvalue $\lambda = 1$, then the left eigenvector normalized to sum to 1 gives the long-term prob for each state (proportion of time process is in that state), which agrees with the rows of $\lim_{n \to \infty} P^n$.